

Preliminary Assessment of the Comparability of Census 2000 Long Form Estimates with Census 2000 Supplementary Survey Estimates

Guidance for Understanding Differences Between Census 2000 Demographic Profiles and Previously Released Census 2000 Supplementary Survey Estimates

Once estimates are released from the Census 2000 long form in the Spring and Summer of 2002, *the long form is the official and best source of estimates for small geographic areas for 2000*. The Census 2000 Supplementary Survey (C2SS), whose results were first released in July 2001, is no longer recommended as the source of such estimates. The C2SS was part of the demonstration program for the American Community Survey (ACS). Its primary objective was to evaluate the feasibility of collecting long form data outside the decennial census. In addition, the data allow us to understand the impact of differences in ACS collection methods compared with the long form methods. Specifically, its results were a “first look” at long form-type estimates, and its results will help users understand how estimates from the ACS will differ from the long form. C2SS estimates will also allow the examination of changes between 2000 and 2001 for areas of 1 million or more, the latter year’s information being obtained from the 2001 Supplemental Survey (SS01). Further, since C2SS population and housing unit counts are controlled to Census 2000 counts of the population and housing units (all sample surveys use weighting to match population controls), it does not give an independent estimate of the population in households.

Two ACS evaluation reports have been completed – on operational feasibility and survey quality. Four more ACS evaluation reports comparing Census 2000 results with C2SS estimates are planned. Report #3 will compare short-form Census 2000 data with C2SS demographic characteristics (to be released in August 2002). Long form data comparisons will be issued from October 2002-January 2003, focusing on social characteristics (#5), housing characteristics (#6), and economic characteristics (#7).

I. Background and General Advice

The key to understanding the differences between estimates from the Census 2000 long form and the C2SS is to remember a truism from the science of survey methodology: “Different surveys using different methods will yield different results.” This is true almost regardless of the extent of efforts to maintain comparability. That does not mean that one or the other survey is flawed. If the surveys are of good quality (and certainly C2SS and the long form qualify in that regard), it is important to recognize, and then to try to explain and understand those differences.

Different surveys also serve different purposes, and it is good to recognize that those different purposes may well underlie the different methodologies used to collect and process the data. In addition, the C2SS uses methods designed to take advantage of the continuous collection that is inherent in the ACS approach. We discuss the major survey differences below and how they might affect interpretation of the surveys’ estimates.

Sampling Error. All surveys have sampling error. Sampling error occurs because only part of the population is directly contacted. With any sample, differences are likely to exist between the characteristics of the sampled units and the entire population from which the sample was chosen. Larger surveys have lower sampling error. The C2SS, while the largest household survey ever

done by the Census Bureau outside of a decennial census, with 700,000 housing units in the sample, is still small when compared to the 19 million housing units selected for the long form. The estimates in the two surveys could well differ because of sampling error, which will be much lower for the long form estimates than for the C2SS estimates for the same geographic area. The Census Bureau provides 90 percent confidence intervals with survey estimates, and these can be used to assess whether apparently different point estimates from different surveys are statistically different. Note that 5-year average ACS estimates will be based on 15 million housing units.

Design and Implementation. The C2SS design and implementation is a dramatic departure from the traditional decennial long form survey. C2SS, as a demonstration program for the ACS, used the basic ACS design, which as an ongoing survey, will continually improve its survey methods to ensure data quality is maintained or enhanced. Given that the primary purpose of the decennial census is to enumerate the U.S. population and housing – not to collect detailed long form data – the long form survey is conducted basically the same way and at the same time as short form data collection. Consequently, the long form may either benefit or be harmed from the decennial census environment. For example, undercoverage (missing people or housing units) is probably less of a problem for the long form than for other household surveys.

Collecting data from households is a formidable task. Because the long form is part of the decennial census enumeration, Census Bureau managers are dependent on the hiring and training of hundreds of thousands of temporary employees. Due to schedule and budget constraints, these employees undergo limited training, and as only 1 in every 5 nonresponse cases are long form households, there is less focus on long form data collection than if it were the only questionnaire being used.¹ In contrast, C2SS/ACS staff are permanent professional interviewers (though a substantial fraction were new hires in late 1999 to handle the C2SS). These professional interviewers are typically experienced and have undergone extensive training, including methods to encourage participation in the survey. This better level of preparedness is expected to lead to both higher response rates and better interview quality.

In addition, because of schedule and budget concerns, the long form did not include a content followup. That is, if items on questionnaires returned by mail were left blank, no attempt to collect the missing information was made unless the short form items were not answered. Subsequently, these items were statistically inferred. The quality of data items, such as income, that are used to distribute federal funding could be affected. In contrast, following up with respondents to obtain answers to questions is an integral part of the ACS methodology to reduce item nonresponse. Unfortunately, the C2SS telephone follow-up was not as successful as planned because of high demand for telephone questionnaire assistance for Census 2000. Resources were available to improve this operation for the SS01 and SS02.

Data Collection Methodology. Another major contrast between the two approaches is the data collection methodology. The long form data are collected as part of the one-time census enumeration activity. Mailing or personal delivery of 19 million Census 2000 long forms and 98 million short forms took place in mid-March 2000 (the long forms of course obtained short form information as well), except in remote Alaska. Due to schedule and operational constraints, a

¹ Because the mail return rate of long forms is lower than for short forms, a larger fraction (22 percent) of the nonresponse caseload is long forms than its fraction (17 percent) of all housing units selected.

replacement questionnaire (follow-up mailing to non-responding housing units) was not feasible. Follow-up for housing units not returning forms by mail began in late April 2000. Hundreds of thousands of temporary Census Bureau enumerators began streaming out nationwide, armed with paper questionnaires and their respective lists of addresses from which no mailed response had been received. Nonresponse followup lasted about nine weeks. While the snapshot of America was taken as of April 1, 2000, readying the long form data for release to data users occurred from about October 2000 until the spring of 2002, and took second chair to the short form (enumeration) data.

In contrast, the C2SS data are collected in ongoing, three-month cycles through a combination of mail-out/mail-back questionnaires, Computer Assisted Telephone Interviewing (CATI), and Computer Assisted Personal Interviewing (CAPI). A sample of housing unit addresses is drawn from the Master Address File, and questionnaires are mailed just prior to the beginning of month one. A second (targeted) questionnaire is then mailed to those who did not respond to the initial mailing. During month two, housing units that did not respond to either mailing and whose telephone numbers we can identify, are called. Finally, in month three, a 1-in-3 subsample of the remaining nonrespondent housing units is drawn, and a personal visit by professional interviewers is made to those housing units. We expect that the quality of data is improved by using well-trained permanent interviewers with computers to complete interviews. The use of the computerized questionnaire, rather than a paper questionnaire, allows the Census Bureau to incorporate consistency checks of the data into the collection process.

Residence Rules. We expect differences in the vacancy rate between C2SS and Census 2000 because of the difference between the C2SS “annual average estimates for ‘current residents’” and the decennial census “April 1, 2000 number for ‘usual residents’”. For example, if your mother lives by herself for 8 months of the year and with you for 4 months of the year, the census counts this as two households and no vacant units, while the C2SS, depending on when the housing units was in sample, might include only one household, with her housing unit as vacant for the part of the year when she’s living with you. In other living situations, the difference could go the other way. For example, the longer C2SS interview cycle gives a vacant unit much more of an opportunity to be occupied than the long form fixed reference date interview method. If a decennial census enumerator shows up at an occupied nonresponse housing unit and is told it was unoccupied on Census Day, it is considered a vacant unit. In contrast, if a C2SS field representative shows up and the unit is occupied, it is counted as an occupied unit regardless of whether it was occupied during the mail-out month (two months earlier). In general, the greatest differences are expected in “highly seasonal” areas. Census 2000 required confirmation of all vacant units whereas one visit by a C2SS interviewer could be sufficient to identify a vacant unit.

In a decennial census, each person is to be enumerated as an inhabitant of his or her "usual residence." Usual residence is the place where the person lives and sleeps most of the time, or the place he or she considers to be his or her usual residence. This place is not necessarily the same as the person's legal or voting residence, nor is it always the same as the place where they happened to be staying on Census Day. The ACS concept of residence differs from this decennial census concept. Because the survey is conducted every month on independent samples, and produces annual or average annual estimates, a concept of current residence was adopted as being more appropriate than the usual residence concept of the decennial census. The census requires that everyone have only one usual residence, since its primary purpose is to provide

counts for the apportionment of congress. The concept of current residence being used in the C2SS/ACS also requires that each person have only one residence at any point in time, but that residence does not have to be in the same place throughout the year. For housing units, each person included in the survey must be a current resident of the unit in sample--a place where he or she is staying for more than two months at the time of survey contact. This concept allows the survey information to more closely reflect the actual characteristics of each area.²

In the vast majority of areas of the country, the use of usual residence or current residence as the classification basis would produce substantially the same statistics. However, appreciable differences may occur for areas where large numbers of people spend several months of the year in units that are not their primary residences; such as Florida, Arizona, and beach or mountain vacation areas with large seasonal populations.

Group Quarters. Another major difference between C2SS and the long form is in their treatment of residents of Group Quarters (GQs). The Census Bureau classifies all people not living in housing units as living in group quarters. Housing units are defined as a house, apartment, a mobile home or trailer, a group of rooms or a single room occupied as a separate living quarters or, if vacant, intended for occupancy as a separate living quarters.³ There are two types of group quarters: institutional (for example, correctional facilities, nursing homes, and mental hospitals) and non-institutional (for example, college dormitories, military barracks, group homes, missions, and shelters). Since the C2SS was designed as a demonstration program for the ACS, and the Census Bureau did not want to burden GQs with duplicate data collection during 2000,⁴ GQs and their data were excluded from the C2SS. This exclusion can have major effects on the estimates for certain areas. Take for instance the median age of a population. If the C2SS and the long form estimates for a locality with a large military facility were compared, the C2SS median age would undoubtedly be higher, as the military tends to employ a population with an age distribution much lower than the general population. On the other hand, the median number of rooms of owned dwellings will not be affected by this exclusion, as group quarters are not in the universe for either survey's estimate.

Reference Period. Another source of potential difference is the reference period for the questions asked. A good example of where differences might arise is in the income questions. The long form asked about income for the 1999 calendar year, while C2SS estimates were

² To implement the current residence concept, the "Two-Month" rule was established. This rule states that if a person is staying in a sample unit at the time of survey contact, and is staying there for more than 2 months, he or she is a current resident of that unit whether or not the unit is also the person's usual residence under census rules. If a person who usually lives in the unit is away for more than 2 months at the time of survey contact, he or she is not a current resident of that unit. Anyone staying in the unit at the time of survey contact who has no other place where they usually stay is considered a resident of the unit. The time of survey contact is defined to be when the respondent completes the survey questionnaire, or when the unit is reached by telephone or through a personal visit during the followup for mail nonresponse.

³ Separate living quarters are those in which the occupants live separately from any other individuals in the building and which have direct access from outside the building or through a common hall.

⁴ Data collection at GQs sometimes takes place using the facility's administrative records, particularly in institutions where the residents cannot be personally interviewed, but always involves the cooperation of the facility's staff.

collected on a rolling basis every month throughout calendar year 2000 and the questions were asked about income “in the previous 12 months.” For example, those interviewed in January 2000 were asked about the January to December 1999 period, and those interviewed in December 2000 were asked about the December 1999 to November 2000 period. Thus, the calendar time period for C2SS income estimates does not correspond to that for Census 2000 (the previous calendar year).

Simple inflation adjustment cannot adjust for the difference in reference periods. We know from research there are reporting differences as a result of the different reference periods (the “previous 12-month” reference period produces lower median wage estimates than the “previous calendar year” reference period),⁵ and there may well be real economic phenomena affecting the levels of income differentially. Obviously, anything affecting income will also affect poverty measurement.

Similarly, other data items can be affected by the reference period difference. Take for example commuting time. If a new bridge or highway were to open in July 2000 that substantially reduced commuting time in one metropolitan area, the estimates of travel time from the C2SS would be affected, while the estimates from the long form would not.

Survey completeness. Missed households, people, and housing units can have critical effects on survey estimates and can be measured directly through comparisons to other aggregate estimates that serve as independent benchmarks or through case-by-case matching studies. When another aggregate estimate of the study population is available (for example, from the decennial census or demographic projections), the level of survey completeness is commonly measured using a ratio calculated as the population estimated by the survey (before applying any adjustment for nonresponse) divided by the population from an independent benchmark. Census Bureau surveys often use the decennial census or the subsequent intercensal population estimates as the aggregate estimate.

To evaluate the C2SS, sample completeness measures were calculated relative to Census 2000. These measures of sample completeness indicate how representative the sample is overall, and for various population groups, relative to the decennial counts. A similar ratio was calculated for the 1990 long form relative to the 1990 census (estimates for the Census 2000 decennial long form are not yet available). As shown in American Community Survey Research and Evaluation Report #2,⁶ the C2SS sample completeness ratios for the total population and for selected racial and ethnic groups were comparable to the ratios for the 1990 decennial long form.

Race and Ethnicity. There are significant differences in the races reported by Hispanic respondents, and some differences in the two-or-more/one-race results, between the C2SS and

⁵ Kirby G. Posey and Edward Welniak, “Income in the ACS: Comparisons to the 1990 Census,” prepared for the March 1998 ACS Symposium, revised June 1999; see <http://www.census.gov/acs/www/Library/Papers/Paper13.htm>.

⁶ “Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey: Report 2: Demonstrating Survey Quality,” forthcoming.

Census 2000.⁷ The concept of race is complex and it appears that very minor differences in how the data are collected can affect the responses to a far greater degree than previously understood. This situation has resulted in plans for extensive testing of race questions for future use in the ACS and the 2010 Census. These differences might affect the estimates of other characteristics, but only to the extent that characteristics associated with how people report their race or ethnicity are correlated with those characteristics.

Results are preliminary. As noted above, the Census Bureau is undertaking an evaluation of differences between the C2SS and the long form, and an evaluation of the differences between the long form and the 3-year average estimates in the 36 ACS test counties is planned, once data processing is completed for the 2001 data. The purpose of those evaluations is to help the user understand how the estimates will differ, but not to adjust the C2SS in any way to mirror the long form. The ACS is intended to provide long form-type data to data users on a more timely basis than the census, and to replace the long form for the 2010 Census. Users need to understand the differences in order to properly use the C2SS and ACS data in their own applications and to be able to distinguish real changes over time from changes in estimates because of differences in methods.

Guidance on comparisons. The table below gives our best advice about when we expect estimates to be comparable, when differences are expected in certain areas, and when the estimates are likely to be noticeably affected by differences in methods.

⁷ Compared with Census 2000, C2SS tends to estimate a larger number of Whites alone, Asians alone, and Native Hawaiians and Other Pacific Islanders alone; and estimate a smaller number of Blacks or African Americans alone, American Indians and Alaska Natives alone, Some other race, and Two or more races population. These results may be related to the reporting of race by Hispanics. The pattern for the race alone or in combination populations was similar to the race alone populations with two notable exceptions. First, the C2SS estimate for American Indian and Alaska Native population alone or in combination population is larger. Second, the Native Hawaiian and Other Pacific Islander alone or in combination population is smaller. When C2SS data on race are compared with Census 2000 data, the percent distribution of C2SS shows that respondents identified themselves about 1.8 percent more often as White alone not Hispanic, counterbalanced by a drop of about 1.6 percent in those who self-identified as Hispanic, "Some other race" alone. There were no other significant differences in the C2SS race distributions by race and Hispanic origin when compared with comparable Census 2000 race distributions.

Understanding why estimates of particular profile characteristics may differ between the Census 2000 long form and the Census 2000 Supplementary Survey

Characteristic	Likely to be unaffected by methodological differences	Probably affected by methodological differences	Affected substantially by methodological differences
General Demographic Characteristics (Demographic Profile: DP-1)			
Sex		GQ	
	There are more elderly women in institutions like nursing homes than elderly men; military bases tend to contain more young men than young women.		
Age		GQ	
	There are more elderly women in institutions like nursing homes than elderly men; military bases tend to contain a younger population than average.		
Race			Q'aies
	The race questions differed.		
Ethnicity (Hispanic)			Q'aies
	The ethnicity questions differed.		
Relationship		DP(in households)	GQ(in group quarters)
	The editing of household relationship differed; relationship is typically not collected in group quarters.		
Households by Type		GQ	DP
	Relationship is typically not collected in group quarters; see also relationship for household data.		
Vacancy Rates			DP
	Vacant units had more time to be occupied in C2SS (see text).		
Homeownership Rate	X		
Household Size	X		
Selected Social Characteristics (Demographic Profile: DP-2)			
School Enrollment		RP	
	Data collection takes place at different times of the year.		
Educational Attainment		GQ	
	Elderly are over-represented in institutions and they tend to have lower educational attainment.		
Marital Status		GQ	
	Relationship is typically not collected in group quarters.		
Grandparents as caregivers	X		

Characteristic	Likely to be unaffected by methodological differences	Probably affected by methodological differences	Affected substantially by methodological differences
Veteran Status		GQ	
	Elderly veterans, particularly of World War II and the Korean War, are over-represented in nursing homes compared with the general population, but so are elderly women who are not as likely to be veterans.		
Disability Status			Q'aies
	The disability questions differed.		
Residence in 1995/1999			Q'aies, RP
	Besides the reference period difference, the long form asked about residence in 1995 while the C2SS asked about 1999.		
Nativity/Foreign Born	X		
Language Spoken at Home	X		
Ancestry	X		
Selected Economic Characteristics (Demographic Profile: DP-3)			
Employment Status		GQ, YR	
	Data are collected all year on C2SS; group quarters residents are less frequently employed.		
Commuting to Work		GQ, YR	
	Data are collected all year on C2SS; group quarters residents are less frequently employed.		
Occupation	X		
Industry	X		
Class [Type] of Worker	X		
Income in 1999		RP, YR (all measures) GQ (per capita)	DP(median earnings)
	Data are collected all year on C2SS; the reference period for income differs; group quarters residents are infrequently employed and tend to have lower incomes; C2SS adjusts dollar amounts to 2000 dollars while the long form reports 1999 dollars; median earnings in C2SS are based on all people with earnings but long form estimates include only people who worked full-time year-round.		
Poverty in 1999		GQ, YR	
	See comments on income.		
Selected Housing Characteristics (Demographic Profile: DP-4)			
Units in Structure	X		
Year Structure Built		YR	

Characteristic	Likely to be unaffected by methodological differences	Probably affected by methodological differences	Affected substantially by methodological differences
	Data are collected all year on C2SS; C2SS will include more homes built in 2000.		
Rooms	X		
Year Moved into Unit		YR	
	Data are collected all year on C2SS.		
Vehicles Available	X		
House Heating Fuel	X		
Lacking Selected Facilities	X		
Occupants per Room	X		
Value		YR	
	Data are collected all year on C2SS.		
Mortgage Status	X		
Monthly Owner Costs		YR	
	Data are collected all year on C2SS		
Gross Rent; Gross Rent as % of Income		YR	
	The reference period differs.		

Notes:

DP=data collection, processing, edit, and/or reporting procedures differed between the C2SS and the long form.

GQ=characteristic affected by group quarters exclusion from C2SS.

Q'aires=question wording differed between the C2SS and the long form.

RP=characteristic has different reference period in long form and C2SS.

YR=data are collected year round on C2SS, at one point in time on long form.